DOCUMENT RESUME

ED 136 316 CS 501 641

AUTHOR Smith, Raymond G.
TITLE Development and Possibilities of Message Measurement Inventories.
PUB DATE Apr 77
NOTE 6p.; Paper presented at the Annual Meeting of the Central States Speech Association (Southfield, Michigan, April 14-16, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Communication (Thought Transfer); *Evaluation; *Measurement Instruments; *Measurement Techniques; Research; *Speech Communication; Test Construction; Test Reliability; Test Validity
IDENTIFIERS *Message Measurement Inventory

ABSTRACT

The Message Measurement Inventory was designed to determine the dimensions which listeners consider when they judge a message. This paper outlines the development of the inventory and describes some of the first studies using it. In addition, the paper discusses tests of the validity, reliability, and precision of the scales and scaling procedures of the instrument. Six general observations, made from the tests of the instrument and the applications made to date, are presented. (JM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED136316

DEVELOPMENT AND POSSIBILITIES OF MESSAGE MEASUREMENT INVENTORIES*

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Raymond G. Smith
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Raymond G. Smith

Indiana University

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Most of us have spent substantial portions of our lives in analyzing and describing effective communications and trying to teach others to do the same. After attempting to cope with problems of criticism more years than I care to acknowledge, I finally began to focus on the listener rather than the speaker. This is an acknowledgment of the primitive notion that the reality of communication, like beauty, rests in the eye and ear of the beholder. In the transactional process the speaker serves as the tran--it is in the listener that the action takes place.

The following brief statement constitutes a summary of a 260 page report describing these efforts. About five years ago a very perceptive student in one of my graduate courses asked me to identify the dimensions of affective appeal in any communication. I responded that the identification of such dimensions would constitute an excellent graduate project. Further contemplation of the question prompted me to extend it to other modes of support as well. So the question in broader form becomes, "What are the perceptual dimensions of a communication?" "When listeners judge messages, what dimensions do they take into consideration?"

The most obvious means of attempting to answer a question of this scope is the approach taken by Charles Osgood and associates when they attempted to identify the dimensions of connotative meaning. So I began to collect all of the qualifying terms that speech critics, educators, psychologists, social scientists, political scientists, historians, and others have variously applied in describing and evaluating communications. Many of these are well known to us in making critiques of student speeches. They include such terms as clear, honest, effective, persuasive, logical, clever, skillful and the like. From a list of several thousand such terms, all but about 500 were eliminated as being esoteric, ambiguous, repetitious.

The first step was to learn which of these terms were generally meaningful to the college population. The factor analytic approach employed by Osgood seemed to be the obvious means for answering this question. But 500 terms is too many for a subject to judge, so I arbitrarily categorized them into four sets. First, those concerned with message thought and content which were labeled rational, those concerned with message emotion, labelled affective, those applicable to credibility, and those concerned with the esthetic elements of the message, labeled artistry.

To further reduce the rating task, the terms were separated into positive and negative sets for each of the four categories, with an effort made to include polar opposites in each group. Following Cattell's advice, marker or probe terms were included in each of the eight sets of terms. These probe terms were taken from Osgood's basic factor structure, the evaluative, activity, and potency categories.

Whenever it was unclear into which category a term was to be placed, it was

*Paper delivered at Central States Speech Association Conference, April 14, 1977.

CS 501 641

included in two or more. The polarities of some terms were in doubt, so these were included in both positive and negative categories.

The data consisted of scale ratings along a 10-point scale in response to the question, "How important is this attribute (such as clarity) to any message? Since the instrument was initially designed for research purposes with college students as subjects, cross sections of the college population, ranging from those taking their first speech course, to advanced doctoral students served as raters. About 800 raters supplied the data. Various factor analytic programs which will be described later, were employed in the analysis. The question to be answered at this time was what constitutes the total map or public common-value-system used by college students in making judgments of communications?

Terms were initially retained from the factor analyses that met a minimal principal dimension loading of .50 and were minimally contaminated (.30) with any other factor. In the final choice, terms were required to meet a loading criterion of .60, and most were well above that criterion. The final selection included 114 scales, 56 positive and 58 negative. From 1 to 4 scales made up each of the 31 positive and 29 negative factors that emerged as the principal dimensions of communication. It is believed that these 60 factors constitute most of the dimensions along which listeners judge messages. Of course not all dimensions have meaning for all subjects. Each subject simply marks zero for any term that for him does not apply, and the remaining terms then constitute his evaluative instrument. Thus, in a sense, each judge creates his own scale by constructing his private rating instrument from the total public meaning space offered by the message measurement inventory.

In applying the instrument in judging a message, two ratings are obtained for each scale. First is a general rating of each scale. The judge is asked to indicate how important a particular scale is for him in the judgment of any message. If the trait is clarity, we ask him to indicate on a scale running from zero to 9 how important clarity is to any message. This is a subjective judgment. Then we ask him to judge a particular message. How clear is this message? This is an objective judgment. The rating for each of the 114 scales is then obtained as the geometric mean of these two judgments. Factor ratings are computed as averages of their component scale judgments.

The instrument yields, in addition to the 60 factor indexes, two average ratings for the total message, one positive and one negative. It also yields two ratings for the so-called rational, affective, credibility, and the artistic factors. A computer program has been written to do all of this hard work.

Much of the work in instrument development was done by my colleagues and friends. Sixteen of my colleagues at Indiana University provided class time and student subjects for the study. I was given expert statistical, design, and computer advice by colleagues in other departments. Also, several friends from other universities provided time and subjects, including Paul Brandes of North Carolina; Al Goldberg of Denver; Ken Frandsen, Penn State; Ed Robinson, Ohio Wesleyan; Jack Whitehead, Texas; and Gordon Wiseman of Ohio University.

Although I am certain that he would disclaim me, Charles Osgood of Illinois provided much friendly advice and encouragement during the late 1950's when I was trying to understand the semantic differential, and Norman Anderson of the University of California, San Diego supplied the conceptual framework for the measurement profile.

Initial tests of the scales and their method of application were conducted by graduate students as graduate class and dissertation projects. The first test was that of the positive objective scales and was made by Judy Pearson as a research project. She randomly distributed eight paragraphs from each of two speeches by Lester Maddox to 322 subjects and had them rated on the 56 positive scales. The eight sets of ratings for each speech were pooled for analysis. The results showed 26 significant differences between the two speeches with all but 5 significant beyond the .01 level.

The first study in which both subjective and objective scale ratings were collected was done as a doctoral dissertation by Lowell Lynn in 1974. He compared two forms of a message on the positive scales, one form he termed subjective, and one objective, and attributed them to three different sources. Using the computer program he condensed the positive scales and ran tests over the 31 factors. He found 17 significant differences between the two messages when attributed to an anonymous source, 11 with the professor as source, and 4 with student as source--32 significant differences out of a total of 93 comparisons.

The initial test of the full 114 scales and their 60 factors was carried out by Tom Clark and myself. But I will let Tom tell you all about that. We have collected for this program a number of papers by colleagues who have used the MMI for experimental or descriptive studies.

In addition to the research reports that you are about to hear, a study has been completed of the Carter-Ford debates. This study is scheduled to appear in a forthcoming issue of the Central States Speech Journal.

Regis O'Connor of Western Kentucky University and I have analyzed most of the data from a study designed to compare source credibility, message credibility coupled with source credibility, and message credibility apart from source.

A most recent study can only be described as a "fun" study. Moya Andrews and I have just completed data collection for comparing the perceptions of three male voices reading a prose selection when the readers were sober, compared to the same three readers reading the same selection when they were something less than sober. (The outcome from this study should assure us both a place on the program of the 1978 CSSA Convention!)

Sue DeWine and I are doing a couple of studies of group credibility, one of participant attribution, and one of opinion change resulting from Devil's Advocacy.

So you have some notion of the nature and scope of the work that is underway with the MMI. All of this work would, of course, be at best questionable and at worst worthless if the instrument is not reliable and valid. I have prepared a section of this report to answer questions of validity and reliability, and will be happy to present it if there is any interest. But I believe this introduction has occupied your attention for long enough and that you should now hear from the other panelists. Mr. Chairman, I desire to yield at this time.

MMI VALIDITY, RELIABILITY AND PRECISION

All of the studies that have been described to you would be trivial or worse if the scales and scaling procedures lack validity, reliability and sensitivity. Therefore series of tests extending over a period of two years have been made. The following paragraphs briefly describe some of these.

First, as previously mentioned, a series of factor analyses, 19 in all, were run on subject ratings of the importances to the subjects of the various scales describing communication. The final two of these involved data from eight separate studies.

To test for factor invariance across factor analyses, two different orthogonal programs were used. To test for factor invariance across method of rotation, both orthogonal and oblique programs (BMD) were run for each of the eight basic analyses. Each factor analysis extracted either 9 or 10 factors which accounted for approximately 60% of the total variance.

Scale order effect and context effect were tested by repeating scales for the same subjects. As many as eight estimates of means and standard deviations for the same scales were thus obtained. One scale, namely "calm" was included among both positive and negative terms, and, interestingly enough, factored with both.

Regression programs were run to check the predictability of the four categories of terms. An elaborate and time-consuming test was constructed to determine whether the categories, namely the rational, affective, credible, and artistic are overlapping. A multivariate analysis was run to learn whether all 60 factors which were presumably orthogonal were in fact so intercorrelated that the 60 scores could be viewed as measures of the same thing. A discriminant function analysis was made for the scores for three widely differing types of messages with no clear pattern emerging across treatments for any factor, suggesting that the discriminant power of each is a function of the particular application. That is, the factors do function independently for various treatment/message conditions.

The Kuder-Richardson reliability formula was applied to test the reliability of both subjects and scales. Also a test-retest correlation of reliability was run for each of the 114 scales making up the final instrument. In addition a fairly comprehensive reliability test was run on data from two sets of 7 messages each. In this test, all except 8 of the 120 comparisons provided indexes significant beyond the .01 level of confidence.

The question of content validity was met by providing as wide as possible a selection of message qualifiers. In the perceptual area this appears to be about the only approach available. The question of construct validity was met through the expedient of selecting as many qualifiers as could be found from those actually used by communication critics engaged in judging messages. Kerlinger has pointed out that, "Whenever hypotheses are tested, whenever relations are empirically studied, construct validity is involved." and, "Factor analysis is perhaps the most powerful method of construct validation."

To test criterion validity for each of the 60 factors, the instrument was applied to two patently different sorts of messages--an informative lecture on the topic of "listening" and the oral reading of a passage from the test

passage, "Androcles and the Lion." Seven of the 60 contrasts failed to show significant differences beyond the .05 level, but most were beyond the .005 level. It is possible that the two messages were in fact no different on the dimensions of candidness, theatricality, cunningness, uncooperativeness, bias, boastfulness and passivity.

CONCLUSION

Several general observations may be made from these tests of the instrument and the applications that have been made to date.

First, many changes occur within listeners in response to messages. These shifts are not ordinarily revealed by many of the tests that have previously been applied. In fact, our data reveal that shifts on individual dimensions sometimes if not frequently cancel each other, in which case the more molar type response measure would show nothing.

Second, extreme caution must be observed by any experimenter in assigning rational, affective, credibility or artistic roles to various sections of a message. A normative agreement exists, but large variability exists both between subjects and within the same subjects at different times or for different contexts.

Third, a remarkable stability exists across both contexts and groups of judges for the importance ratings of various communication dimensions, although as noted above, great variation exists between judges.

Fourth, Osgood's evaluative, activity, and potency factors, ubiquitous in many other types of judgments, enter in to and are related to judgments of messages, but less closely than in the judgments of other classes of objects.

Fifth, the bipolarity assumed by Osgood in his explorations of the dimensions of meaning does not apply to the ratings of a communication, at least not universally. It appears that many scales are bipolar, but many are not.

Sixth, the categories of message influence--rational, affective, credible, and artistic are sufficiently discrete to be helpful in analysis. In groups of messages they appear in the following order: Rational, credible, affective, and artistic. They do, however, change orders for particular messages.

Finally, it probably should be noted that an over all factor analysis was run for the 60 factors for 325 subjects. The most important factor to emerge was polarity, with nearly all of the 29 negative factors appearing together. When the 31 positive and 29 negative factors were run separately, the results made a bit more sense. There appear to be six undergirding dimensions in communication, three positive and three negative. Two of these--likability and dislikability are orthographically bipolar. The remaining two that add to message effectiveness are creativity and analytical quality; the two that detract are disorderliness and indirectness. As a final caveat, however, whether these last analyses have anything helpful to offer to students of communication is still moot.